# DESIGNING DEEP LEARNING MODEL FOR AUTOMATIC IDENTIFICATION OF CROP WILD RELATIVES

## SHORT SYNOPSIS

*Submitted in partial fulfillment of the requirement of the degree of*

## DOCTOR OF PHILOSOPHY

*to*

*Manav Rachna International Institute of Research and Studies*
*(Deemed to be University)*

by

**Ankur Tomar**
Registration Number: 22/PHD/S/043

*Under the Supervision of*

**Supervisor**
**Dr. Suresh Kumar**
**Professor,**
**CSE Department, SET**

**Co-Supervisor**
**Dr. Kuldeep Tripathi**
**Senior Scientist,**
**ICAR-NBPGR, New Delhi**

**Department of Computer Science & Engineering**
**School of Engineering & Technology**
**Manav Rachna International Institute of Research & Studies**
**Sector 43, Faridabad, India**.

## SCHOOL OF ENGINEERING & TECHNOLOGY

### Certificate

I hereby declare that the short synopsis titled "Designing Deep Learning Model for Automatic Identification of Crop Wild Relatives" being submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Computer Science and Engineering under School of Engineering & Technology of Manav Rachna International Institute of Research & Studies, Faridabad during the academic year 2024-2025.

It is a bona fide record of original work carried out under the supervision of Dr Suresh Kumar and co-supervision of Dr. Kuldeep Tripathi and has not been submitted in full or part to any other university or institute for degree or diploma. Further, it is certified that short synopsis is plagiarism free and proof has been submitted.

(Ankur Tomar)
22/PHD/S/043

04/09/2025

Supervisor
Dr. Suresh Kumar
Professor, Computer Science and Engineering
School of Engineering & Technology

4/9/25

Co-supervisor
Dr. Kuldeep Tripathi
Senior Scientist
ICAR-NBPGR, New Delhi

**(Chairperson, DRC)**

**Dean, SET**

# Abstract

Herbarium specimens contain a vast wealth of genetic resources. Among these resources are the Crop Wild Relatives (CWRs) which play a crucial role in improving crops and preserving genetic diversity. Identifying CWRs automatically from live plant and herbarium images is a complex challenge due to the variations in plant shapes and the need for accurate species recognition. Convolutional Neural Networks (CNNs) have proven to be highly effective tools for classifying images, offering promising possibilities for automating CWR identification. This concise summary delves into CNN models specifically designed for plant identification across various formats, including herbarium specimens. The underlying design principles are explored, performance evaluated, and practical applications highlighted in safeguarding biodiversity and advancing agricultural research.

In recent years, researchers have made significant advancements in the development of Convolutional Neural Networks (CNNs) tailored for CWR identification in herbarium specimens as well as live plant. These CNN models utilize innovative design principles and advanced algorithms to tackle the complexity of plant shapes and ensure accurate species recognition. Through rigorous evaluation, these models have demonstrated exceptional performance in automating the identification process, revolutionizing the conservation of genetic diversity and agricultural research. By capitalizing on the vast genetic resources stored in herbarium specimens, CNN models are empowering scientists to safeguard biodiversity and accelerate breakthroughs in crop improvement.

With the guidance of our supervisor and co-supervisor, we have identified six genera comprising multiple species of Crop Wild Relatives (CWR) for our research. These particular species are infrequently found in their natural habitats, but their cultivated counterparts are well-known among researchers and students alike for their study. The outcome of this research is the development of an automated system leveraging Convolutional Neural Networks (CNNs) for the easy identification of Crop Wild Relatives (CWRs) from herbarium and live plant images. This system simplifies the process for new researchers and taxonomists by providing an efficient and accurate tool for species recognition, thereby facilitating biodiversity studies and genetic resource conservation efforts.



**Keywords**: Artificial Intelligence; CWR; Deep Learning; DenseNet; Herbarium Species; InceptionV3; Machine Learning; MobileNet; Plants Identification; ResNet50; VGG16.

# Table of Contents


| | | |
|---|---|---|
| *Certificate* | | *i* |
| *Abstract* | | *ii* |
| *Table of Contents* | | *iii* |

| **S. No.** | **Content** | **Page No.** |
|---|---|---|
| **1** | Introduction to the Broad Area | 1-5 |
| **2** | Literature Review | 6-13 |
| **3** | Research Gaps | 14 |
| **4** | Problem Statement | 15 |
| **5** | Research Objectives | 16 |
| **7** | Research Methodology | 17-18 |
| **8** | Expected Outcome of the Research | 19 |
| **9** | References | 20-21 |
| **10** | Similarity Verification Report | |
| **11** | Grammarly Certificate | |

## 1. Introduction to the Broad Area

Agriculture is a major part of India's economy, providing food, jobs, and raw materials for many industries. It supports a large share of the population and contributes greatly to national growth [1]. However, the sector faces several challenges in crop production. Declining soil fertility, rising pest attacks, limited water availability, and heavy reliance on a few crop varieties continue to reduce overall productivity [2].

India's climate ranges from dry deserts to humid tropical regions, making agriculture highly complex. These varying conditions often cause unpredictable weather, including droughts, floods, and irregular rainfall, which disrupt traditional farming and reduce crop yields. Such challenges threaten farmers' livelihoods and the nation's food security. Addressing these problems is vital for the long-term stability and growth of the agricultural sector [3]. Crop Wild Relatives (CWR) provide an important solution. These are wild plant species genetically related to cultivated crops and possess useful traits such as drought tolerance, pest resistance, and adaptability to different climates. By incorporating these traits into modern crops, CWRs can help improve resilience, productivity, and food security under changing environmental conditions [4].

India has a rich diversity of Crop Wild Relatives (CWRs), including species related to major food crops. These wild plants are valuable for improving crop quality and maintaining genetic diversity. When cultivated crops are bred with CWRs, they gain better resilience, productivity, and adaptability [5]. This process, known as crop wild relative utilization, helps develop improved crop varieties with stronger traits.

CWRs also play a key role in maintaining genetic diversity within crops. Because wild populations adapt to different environments, they accumulate useful genetic variations over time. Preserving these wild species protects valuable traits that can be used for future crop breeding. Beyond agriculture, CWRs also help sustain ecosystems by supporting pollinators, providing habitats, and serving as food for wildlife. Safeguarding CWR populations and their natural habitats is therefore essential for protecting biodiversity and ensuring long-term environmental and agricultural stability [6].

Key crop genera, including Allium, Lathyrus, Solanum, Triticum, Vigna, and Oryza, were selected for their significant value to researchers. The diversity observed within the Lathyrus genus, as illustrated in Fig. 1, highlights the variation among Crop Wild Relatives (CWRs), underscoring their importance in genetic studies and conservation efforts. These genera provide foundational models for advancing research and innovation in the agricultural field. Many of these crops fall under international mandates, with various research organizations focusing on specific genera to address global food security challenges. For example, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) [7] in India focuses on pulses. The International Center for Agricultural Research in the Dry Areas (ICARDA) [8] in Morocco works on Lathyrus. The Wild Lathyrus accessions are shown in Figure1. Further, International Maize and Wheat Improvement Center(CIMMYT) [9] in Mexico specializes in wheat research (Triticum). These centers, along with others in the Consultative Group on International Agricultural Research (CGIAR) [10] network, play a critical role in improving crop yields, enhancing climate resilience, and promoting sustainable agricultural practices. Collectively, 12 UN-affiliated research institutions are dedicated to advancing agricultural science and ensuring global food security. Their collaborative efforts form the basis for sustainable agricultural development worldwide.

Figure 1: Wild Lathyrus accessions

The Indian Council of Agricultural Research-National Bureau of Plant Genetic Resources (ICAR-NBPGR) and the National Herbarium of Cultivated Plants (NHCP) play a pivotal role in the accurate identification of Crop Wild Relatives (CWR). While manual identification relies on the expertise of taxonomists, challenges arise from sample degradation during prolonged storage. The

expertise of these institutions ensures precise classification, provides authenticated genetic resources, and delivers high-quality reference images critical for developing reliable image classification models. This accuracy minimizes errors, aligning model predictions with botanical standards, and enhances the model's capability to accurately classify CWR, thereby supporting agricultural research and conservation initiatives.

Moreover, herbaria serve as essential resources in this context. Herbaria [11] organize plants by their classification to help find and study them. Each plant is labeled with its name, when and where it was collected, and who collected it. This helps researchers compare and analyze plant species, learn about biodiversity and evolution, discover new species, and track changes in plant distribution over time. Herbaria preserve dried and pressed plants, which ensures future generations can benefit from this botanical knowledge.

The NHCP in New Delhi, India is one of 25 major herbaria in the country. It holds around 26,000 herbarium specimens, showcasing 270 families, 1,500 genera, and 4,156 species [11]. The NHCP also maintains a virtual herbarium with approximately 6,000 images as shown in Figure 2. In addition to plants, the collection includes seeds, fruit samples, and economic products. The NHCP is unique because it represents various forms of crop plants, making it an important resource for studying crop genepools. It features over 500 crop taxa and more than 550 crop wild relatives [12].

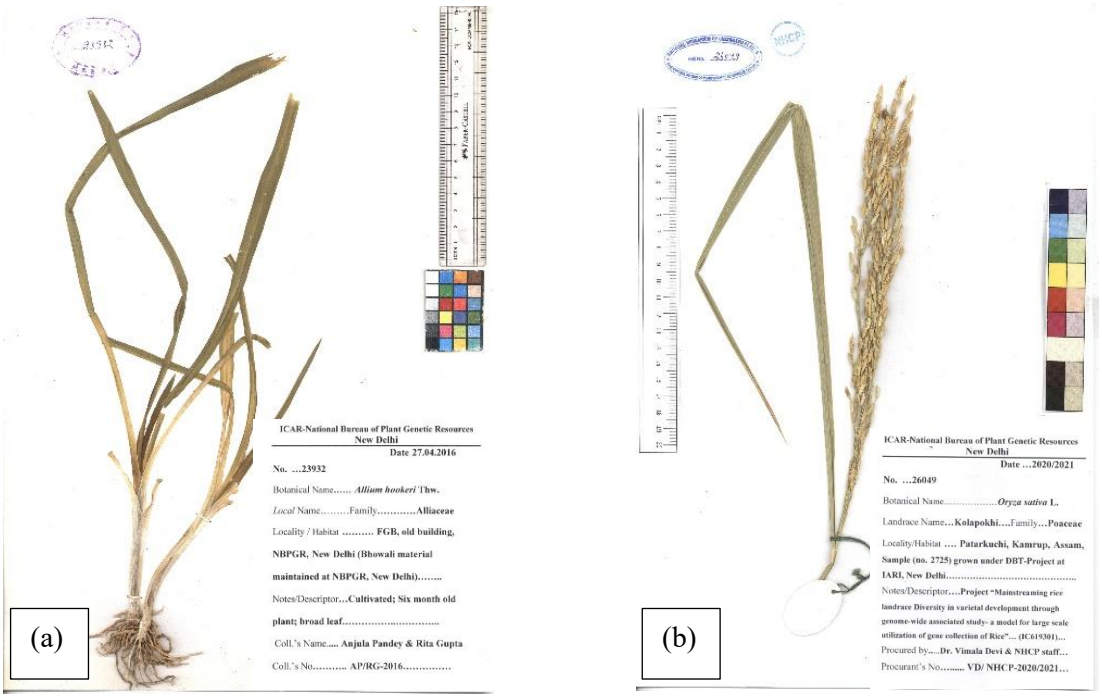
ICAR-National Bureau of Plant Genetic Resources
New Delhi
Date 27.04.2016
No. ...23932
Botanical Name..... Allium hookeri Thw.
Local Name.....Family.....Alliaceae
Locality / Habitat ..... FGB, old building, NBPGR, New Delhi (Bhowali material maintained at NBPGR, New Delhi).....
Notes/Descriptor...Cultivated; Six month old plant; broad leaf.....
Coll.'s Name.... Anjula Pandey & Rita Gupta
Coll.'s No..... AP/RG-2016.....
(a)
ICAR-National Bureau of Plant Genetic Resources
New Delhi
Date ...2020/2021
No. ...26049
Botanical Name.....Oryza sativa L.
Landrace Name...Kolapokhi....Family...Poaceae
Locality/Habitat .... Patarkuchi, Kamrup, Assam, Sample (no. 2725) grown under DBT-Project at IARI, New Delhi.....
Notes/Descriptor....Project "Mainstreaming rice landrace Diversity in varietal development through genome-wide associated study- a model for large scale utilization of gene collection of Rice"... (IC619301)...
Procured by....Dr. Vimala Devi & NHCP staff...
Procurant's No..... VD/ NHCP-2020/2021...
(b)

Figure 2: Herbarium specimens of NHCP used in classification models (a) *Allium hookeri* [36] (b) *Oryza sativa* L. [37]

### 1.1 Artificial Intelligence for CWR Identification

Artificial Intelligence (AI) has emerged as a powerful tool for identifying CWRs. AI technologies can automate CWR identification from plant species databases, including herbarium specimens. This streamlines the process by analyzing plant features [13]. Rapid identification of CWRs is essential for biodiversity conservation and crop improvement, enabling researchers to preserve genetic diversity and develop resilient crop varieties [14]. Advances in deep learning, particularly with Convolutional Neural Networks (CNNs), have transformed the field of plant image identification. CNN-based models accurately classify and identify plant species from images[15]. Convolutional Neural Network (CNN) models offer significant potential for applications in agriculture, biodiversity conservation, and ecological research. By automating plant identification, these models facilitate rapid species assessment, biodiversity monitoring, and crop improvement efforts. Their implementation plays a crucial role in advancing agricultural research and supporting biodiversity conservation initiatives. It streamlines processes by saving time and labor, facilitates the documentation and monitoring of plant biodiversity, supports the conservation of Crop Wild Relatives, aids in detecting invasive species, and generates valuable data for ecological research [16].

### 1.2 Challenges in Crop Wild Relatives (CWR) identification

**Expertise Gap:** The gradual retirement of experienced taxonomists has led to a growing shortage of specialized knowledge in plant identification. Many new researchers lack the depth of expertise needed for precise species classification, particularly within diverse and complex groups such as Crop Wild Relatives (CWRs). This gap poses a challenge for maintaining accuracy and consistency in taxonomic research and conservation efforts.

**Time and Cost Efficiency:** Manual identification of species from herbarium specimens demands considerable time, effort, and specialized expertise. By automating this process, AI-based systems can greatly reduce the time and labor involved, making species identification faster, more efficient, and cost-effective for large-scale botanical studies.

**Scalability:** AI-based identification systems can handle and analyze large numbers of herbarium specimens quickly and accurately. This allows researchers to study extensive datasets that would be difficult or impossible to process manually, improving the scope and efficiency of botanical research.

**Consistency and Standardization:** AI algorithms offer consistent and standardized identification results, minimizing errors and discrepancies that may arise from human subjectivity or variation in expertise.

**Accessibility:** AI-based identification tools can be made accessible online, allowing researchers worldwide to utilize them regardless of their geographic location or access to taxonomic expertise.

**Bridging the Knowledge Gap:** AI-driven systems can act as effective tools for knowledge transfer, supporting new researchers by providing automated species identification capabilities that reflect expert-level accuracy. This approach helps mitigate the impact of retiring taxonomists and preserves valuable taxonomic knowledge. Overall, the integration of AI in Crop Wild Relative (CWR) identification not only compensates for the loss of specialized expertise but also enhances the efficiency, scalability, and accessibility of species identification for researchers and conservationists. Strengthening the identification and conservation of CWRs is crucial for safeguarding agricultural biodiversity and improving crop resilience to climate change, pests, and diseases.

## 2. Literature Review

The literature review focuses on four main areas:

**Herbarium Collections and Their Importance:** An overview of herbarium collections and their key role in preserving plant diversity and supporting taxonomic and ecological research.

**Crop Wild Relatives (CWR):** A discussion on the importance of CWR in agriculture, their contribution to crop improvement, and the difficulties involved in correctly identifying them.

**Deep Learning in Image Analysis:** An introduction to deep learning methods, especially Convolutional Neural Networks (CNNs), and how they are used for image classification and pattern recognition.

**Applications of Deep Learning in Botany:** A review of existing research applying deep learning to plant identification, with emphasis on herbarium specimens and the main challenges reported in these studies.

**Shirai et al.**[17] developed a highly accurate image recognition system to identify taxon names. Using 500,554 specimen images from 2,171 plant taxa in Japan, their system achieved 96.4% accuracy. The research identified specific areas where artificial intelligence systems make decisions and highlighted taxa that are commonly misidentified. This system can be applied globally to select and correct misidentified herbarium specimens, enhancing their utility in research. The generalized ability of this technique is not sufficient and functional at every stage.

**Triki et al.**[18] introduced a deep learning-based approach, to automate the detection and pixel-wise segmentation of leaves using a Mask Region Convolutional Neural Network (Mask R-CNN). Deep Leaf accurately identifies and measures leaf traits such as length, width, perimeter, area, and petiole length from herbarium specimens. Experimental results showcase the efficacy of Deep Leaf in segmenting leaves across various plant families, yielding an average relative error of 4.6% for leaf length and 5.7% for leaf width in comparison to manual measurements by experts. This automated approach streamlines the extraction of morphological traits from herbarium specimens, thereby facilitating research on plant responses to climate change and evolutionary history. However, it should be noted that this method is not suitable for training on image variations and diverse classes such as fruits, herbs, flowers, and non-flowering plants.

**Weaver et al.** [19] developed LeafMachine, a model trained on a dataset of 2,685 herbarium specimens sourced from 138 different herbaria. The model's effectiveness was evaluated using specimens from 20 diverse plant families, encompassing significant variations in resolution, quality, and layout, showcasing its adaptability to diverse data conditions. LeafMachine demonstrated its ability to extract at least one leaf measurement from 82.0% of high-resolution images and 60.8% of low-resolution images. Notably, only 0.9% of the high-resolution images and 2.1% of the low-resolution images that remained unmeasured were deemed to have measurable leaves upon visual inspection. A major limitation highlighted by their study is the substantial time investment needed to process extensive datasets.

**Younis et al.** [20] study showcased the potential of deep learning techniques for detecting and annotating plant organs in digitized herbarium scans, demonstrating their effectiveness in identifying structures such as leaves, stems, and flowers. The authors developed a deep learning model to automate the labeling process, reducing manual effort in specimen analysis. However, several limitations were identified. The dataset lacked sufficient diversity, which restricted the model's ability to generalize across different species and image qualities. The system also struggled with overlapping or complex plant structures, and variations in scan quality further affected accuracy. Additionally, reliance on manual annotations introduced possible labeling errors, while high computational demands limited scalability. Integrating the model into existing herbarium workflows also presented practical challenges, indicating a need for further refinement to improve robustness and usability.

**Chulif et al.** [21] introduced the Herbarium-Field Triplet Loss Network (HFTL network), a model designed to learn relationships between herbarium and field images by reducing intra-species variation and increasing inter-species distinction. In the PlantCLEF 2020 and 2021 evaluations, the HFTL network achieved mean reciprocal rank scores of 0.108 and 0.158, respectively, for species with few available field photographs. The approach demonstrates the potential of combining herbarium and field data to improve species recognition under limited training conditions. However, a key limitation is the relatively high computational and implementation cost associated with such methods.

**Ahlstrand et al.** [22] analyzed four decades of pollen monitoring data to study how 12 allergenic

grass species respond to climate change. The research showed that increasing temperatures have extended pollen seasons and caused earlier flowering in several species. Seven key contributors to airborne pollen were identified based on flowering duration and abundance, with four species exhibiting significant shifts in flowering timing due to rising temperatures. The study highlights the importance of using data from natural history collections to improve pollen forecasting, which is vital for managing asthma and allergy risks under current and future climate conditions. However, a major challenge lies in verifying the phenological stage of grass specimens in herbarium collections, as their small floral structures make accurate assessment difficult.

**Hussein et al.** [23] introduced deep learning techniques to reconstruct damaged herbarium leaves, thereby enhancing the utility of digitized specimens for automated identification. The study compared the performance of two convolutional neural network (CNN) models on data from ten plant families. The use of reconstructed leaf images led to a significant improvement of over 20% in automated identification accuracy compared to using original images with damaged leaves. This underscores the potential of deep learning for reconstructing damaged leaves within herbarium datasets. However, the study also noted that the effectiveness of this approach diminished when subtle variations occurred within species from the same family.

**Little et al.** [24] investigated species identification within the *Melastomataceae* family of flowering plants using a large collection of herbarium specimens—46,469 samples representing 683 species—combined with expert taxonomic input. Despite the uneven distribution of specimens across species, the study formed part of an FGVC (Fine-Grained Visual Categorization) competition that engaged 22 teams submitting 254 models. The top-performing models achieved over 88% accuracy in species recognition. This work highlights how FGVC challenges bridge the fields of botany and computer vision, encouraging collaboration between taxonomists and data scientists. The research specifically focused on data from the *Melastomataceae* family.

**Ott et al.** [25] introduced an extendable pipeline that applies advanced deep learning–based object detection techniques to extract leaf images from herbarium specimens of two *Leucanthemum* species. Using 183 specimens for training, the system accurately detected and extracted intact leaves from 95% of the 61 test specimens. This work established GinJinn as an effective deep learning tool for automatically identifying and segmenting individual leaves or other plant structures from herbarium

sheets. The proposed pipeline offers improved flexibility and accessibility compared to earlier image processing methods that relied on manually designed features.

**Jye et al.** [26] aim to develop an efficient automated system using image processing and pattern recognition techniques to identify three *Ficus* species with closely similar leaf morphology—*F. benjamina*, *F. pellucidopunctata*, and *F. sumatrana*. A total of 54 leaf images were analyzed through three main stages: image preprocessing, feature extraction, and species recognition. Both Artificial Neural Network (ANN) and Support Vector Machine (SVM) models were employed for classification. The system achieved an accuracy of 83.3%, demonstrating its potential for species-level identification within the *Ficus* genus. However, the study was limited to a small dataset encompassing only three *Ficus* species.

**Wäldchen et al.** [27] examined the role of deep learning neural networks in advancing automated species identification. Their study provides an overview of key machine learning frameworks and reviews several deep learning techniques applied to biological image data. The authors also discuss publicly available applications that demonstrate the potential of these methods in taxonomy and ecology. By evaluating models across diverse datasets encompassing plants, animals, and other biological specimens, the study offers a comprehensive view of model accuracy and serves as a foundation for future research in automated species recognition.

**Ballo et al.** [28] aimed to develop a method for recognizing plant species from herbarium specimens using deep learning classification techniques. The researchers tested their approach on a dataset consisting of 7,543 images from ten plant families housed in the national herbarium of Côte d'Ivoire. Several CNN architectures, including InceptionV3, DenseNet-121, VGG19, MobileNet, and ResNet101. In the first experiment, which did not involve preprocessing, model accuracies ranged between 76.94% and 83.47%, with DenseNet-121 achieving the highest accuracy. In contrast, when preprocessing was applied, accuracies improved, ranging from 82.80% to 85.80%, with ResNet121 emerging as the top performer.

The literature review is summarized in Table 1.

**Table 1: Summary of Literature Review**

| Sr.no. | Authors | Objective | Pros | Limitations |
|---|---|---|---|---|
| **1.** | Shirai et al.[17] | An image recognition system was successfully developed to accurately recognize taxon names. | This technique proves valuable for identifying and rectifying misclassified herbarium specimens globallyas it can be applied to digitized photographs from any location. | The generalized ability of this technique is not sufficient. |
| **2.** | Triki et al.[18] | Introduced a deep learning method called Deep Leaf to detect and segment leaves pixel-by-pixel. Based on an enhanced version of the Mask Region Convolutional Neural Network (Mask R-CNN), Deep Leaf accurately identifies and quantifies the morphological characteristics of every leaf in herbarium specimens. | This method is effective to recognize leaves belonging to various families, with an average relative error of 5.7% for leaf width and 4.6% for leaf length compared to manual measurements by professionals in botany and ecology. | This method is not applicable to be trained on image variations and on diverse classes, like fruits, herbs, flowers, and non-flowering plants. |
| **3.** | Weaver et al. [19] | Suggested an application known as LeafMachine for detecting leaves from digitized herbarium specimens or leaf images in an autonomous way. In this, an ensemble of machine learning (ML) models is implemented. | The flexibility of this approach is exhibited for detecting leaves from herbarium specimens by augmenting available trait information | The time consumption is higher on enormous sized data sets. |
| **4.** | Younis, et al. [20] | A suggested deep learning (DL) approach for detecting plant organs on digitized herbarium specimens is based on Region-based Convolutional Neural Network (R-CNN) techniques. | On several herbarium scans for 6 kinds plant organs, the suggested approach works robustly on leaves and stems while detecting plant organs. | This approach is incapable of recognizing flower precisely. |
| **5.** | Chulif et al. [21] | They evaluated the robustness of cross-domain plant characteristics learned through a triplet | The performance of the models was significantly enhanced by utilizing feature | The utilized methods consumed much cost. |

| | | | | |
|---|---|---|---|---|
| | | loss metric learning approach, comparing it to a supervised classification approach. | dictionary generation techniques that incorporated various combinations of herbarium field features. | |
| **6.** | Ahlstrand et al. [22] | presented herbarium specimens to explore the potential impact of single grass species on overall pollen patterns and the variation in flowering times of these species in response to climate changes. | The data taken from herbarium specimens is useful for estimating the pollen season and predicting its shift with respect to future climate change. | The major issue is to verify the phenological phase of grass herbarium specimens as their flowers' size is small. |
| **7.** | Opała-Owczarek et al. [29] | Established a protocol to select and process the historical material from an enormous herbaria with Arctic plants. These resources leads to offer an optimal proxy data source on the varied natural environment | The established protocol exhibited the accuracy of hypothesis of extending dwarf shrub chronologies on the basis of dwarf shrubs available in herbarium cards. | The accuracy of this protocol is mitigated in case of any change in archive of plants. |
| **8.** | Latin et al. [30] | Focused on learning a mapping between herbarium photos and images of leaf plants, and vice versa, through the alignment of two models—one for the leaf domain and another for the herbarium. | Findings from the newly constructed unpaired dataset revealed that a low lambda value for perceptual loss, ranging from 0.025 to 0.05, contributed to reduced fréchet inception distances and higher inception scores. | This models are unable to count plant organs and compute their size. |
| **9.** | Hussein et al.[31] | They concentrated on the application of computer vision and machine learning techniques to digital herbarium specimens, conducting an extensive review of 70 studies. | This study is serve as a valuable resource for professionals in computer science and ecology, offering a robust foundation for future scholars in related fields. | These techniques are not suitable on improper data. |
| **10.** | Hussein et al. [23] | Propose deep learning techniques as a mechanism to regenerate the damaged herbarium leaves for maximizing the worth of the digitized specimens for | The applicability of deep learning approaches for reconstructing damaged leaves even within challenging herbarium | This mechanism offered lower efficiency when subtle variations are occurred in species from a |

| | | | | |
|---|---|---|---|---|
| | | automated plant identification task. | leaf datasets. | single-family. |
| **11.** | Little et al. [24] | Emphasized the potential of machine learning in expediting the development of robust automatic species identification algorithms, thus facilitating the discovery of new species. To generate an automatic species identification method, FGVC6 is executed on Herbarium 2019 dataset. | The study revealed that, for the Melastomataceae family, which includes species represented by varying specimen counts typical of herbarium collections, over 88% of the species were correctly identified by the top four algorithms. | This work concentrates on only the data related to Melastomataceae family. |
| **12.** | Ballo et al. [28] | This study focuses on developing a deep learning-based method for plant species recognition using herbarium specimens. The proposed techniques were tested on a dataset comprising 7,543 images from ten different plant families, sourced from the National Herbarium of Côte d'Ivoire. | It include 10 plant family of national herbarium of Côte d'Ivoire in that include identification process. | This work concentrates on only the data related to only Ivorian plants of herbarium. |
| **13.** | Ott et al. [25] | Introduced a GinJinn as a deep-learning object-detection (DLOD) method to recognize and extract individual leaves or other structures automatically from herbarium specimens. | This method is flexible and offers lower entrance barrier in contrast to other method on the basis of manually generated features. | Strong differences are found in lighting conditions or the angle while capturing the images, used to train and test data. |
| **14.** | Jye et al. [26] | The purpose of the current study is to develop an efficient baseline automated system that uses image processing and pattern recognition techniques to identify three species of Ficus: *F. benjamina*, *F. pellucidopunctata*, and *F.* | A total of 54 leaf image samples were used in this study. The proposed system was developed through three main steps: image pre-processing, feature extraction, and recognition. | This work concentrates on only the data related to only ficus genus which include only three species. |

| | | | | |
|---|---|---|---|---|
| | | *sumatrana*. | | |
| **15.** | Tibbetts et al.[32] | Researchers trained a neural network using random plant samples from a large French herbarium dataset. Subsequently, the network was further trained on a much smaller Costa Rican database. | Despite the differences between French and Costa Rican species, the model achieved 80 percent accuracy in identifying Costa Rican species. | This limits the identification accuracy of French and Costa Rican species to 80 percent. |
| **16.** | Wäldchen et al.[27] | focus on deep learning neural networks as a technology that enabled breakthroughs in automated species identification in the last 2 years. A brief overview of machine learning frameworks applicable to the species identification problem | Include multiple plant, animal species to classify in their categories by using multiple dataset and DL models. | This work concentrated on vast dataset of plants, animals and many different dataset to check the accuracy of models. |
| **17.** | Chulif et al. [33] | They suggested a cross-domain adaptation technique using a model called the Herbarium–Field Triplet Loss Network (HFTL network) to establish a mapping between the herbarium and field domains. | The applicability of this technique is proved to recognize rare species and generalize, even in the absence of field images. | Only 435 species are comprised in the given dataset to compute the efficiency of this technique. |
| **18.** | Kajihara et al.[34] | Determined machine learning methods can serve as effective alternatives for assisting in the recognition of plants within herbarium collections . | The herbarium photos from the Piperaceae family, particularly those of Manekia, Peperomia, and Ottonia, showed potential for identification purposes. | The complexity of some methods is found higher. |
| **19.** | Triki et al.[35] | Identified plant organs in digital herbarium specimen images through the application of a deep learning technique based on the improved YOLO-V3 approach | With precision and recall reaching 94.2% and 95.5%, respectively, demonstrate that our suggested technique exhibits the ability to recognize organs of various sizes within distinct specimens. | The efficacy of this technique is affected due to size of the training dataset and numerous plant portions concerning size, shape, orientation, and color. |

## 3. Research Gaps

Despite significant progress in deep learning and its applications in botany, several research gaps remain unaddressed:

**Limited Annotated Datasets:** One of the primary challenges is the scarcity of comprehensive and annotated datasets of herbarium specimens, particularly those focused on CWR. Existing datasets are often limited in size and diversity, which hampers the development and evaluation of robust deep learning models.

**Unused Dataset of CWR:** A key gap in current research is the limited use of available Crop Wild Relatives (CWR) datasets. Although these datasets are vital for biodiversity conservation and crop improvement, they remain largely unexplored in existing studies. This work addresses that gap by utilizing the CWR dataset for the first time in collaboration with ICAR-NBPGR and NHCP. Through the application of advanced artificial intelligence techniques, the study aims to strengthen species identification, diversity assessment, and conservation planning, ultimately contributing valuable insights to plant genetics, sustainable conservation, and agricultural advancement.

**Model applicability and generalizability:** Although deep learning models have achieved strong results in individual studies, their performance often drops when applied to data from different herbarium collections. Differences in specimen preservation, image quality, lighting, and collection techniques can greatly affect accuracy. This study focuses on developing models that maintain consistent performance across varied datasets, ensuring broader applicability and reliability in real-world botanical and conservation contexts.

**Integration of Metadata:** Metadata linked to herbarium specimens and collected germplasm—such as geographic origin, collection date, and collector details—offers valuable contextual clues for accurate species identification. However, most existing studies have overlooked the potential of combining this information with image data. Incorporating metadata through multimodal learning approaches can strengthen model understanding, leading to improved accuracy and more informed insights in plant classification and biodiversity analysis.

## 4. Problem Statement

Manual identification of Crop Wild Relatives (CWR) remains a major challenge in botanical research and conservation, as it is both time-consuming and prone to human error. This study aims to address this limitation by developing a deep learning model that automates the identification process. The objective is to improve the speed and accuracy of CWR recognition while ensuring that the model remains transparent, interpretable, and adaptable to a wide range of herbarium collections and plant image datasets.

## 5. Research Objectives

To address the identified problem, this study will focus on the following research objectives:

1. **Develop a Comprehensive Annotated Dataset:** Create a detailed dataset of Crop Wild Relatives (CWR) that captures a wide variety of species, image conditions, and herbarium preparation styles. This diversity helps the model perform consistently and accurately across different real-world situations.
2. **Enhance Feature Extraction and Representation Learning:** Explore improved ways to capture and represent visual features using techniques like transfer learning, self-supervised learning, and attention mechanisms. These methods aim to make Crop Wild Relative (CWR) identification more accurate and dependable.
3. **Design and Optimize Deep Learning Models:** Build and fine-tune deep learning models specifically for identifying Crop Wild Relatives (CWR) from herbarium images. Test and compare different model designs to determine which provides the best performance and accuracy.
4. **Implement Real-time Identification Systems:** Develop practical identification tools that work in real time, optimized for mobile and cloud deployment. These systems will support efficient species recognition in the field, aiding conservation and research activities.

## 6. Research Methodology

To achieve the outlined research objectives, this study will be conducted in four key phases as shown in Figure 3:

**Phase 1: Dataset Development and Annotation**

**Objective:** Develop a comprehensive annotated dataset of Crop Wild Relatives (CWR).

1. **Data Collection:** Herbarium images collected from NHCP, ICAR-NBPGR and live samples maintain in institutional field.
2. **Preprocessing & Annotation:** Images are cleaned, standardized, and augmented to improve quality and diversity, then accurately labeled by experts for correct species identification.

**Outcome:** A robust, diverse dataset to improve model generalization.

**Phase 2: Feature Extraction and Representation Learning**

**Objective**: Enhance feature extraction for better CWR identification.

1. **Transfer Learning**: Fine-tune pre-trained models (e.g., ResNet) on the CWR dataset.
2. **Self-Supervised Learning**: Use contrastive learning (e.g., SimCLR) to improve feature representation.
3. **Attention Mechanisms**: Integrate attention layers to focus on relevant plant parts, improving model accuracy.

**Outcome**: Improved feature extraction and model reliability.

**Phase 3: Design and Optimization of Deep Learning Models**

**Objective**: Develop and optimize deep learning models for herbarium image classification.

1. **Model Development**: Build and test various architectures (CNNs, Transformers).
2. **Evaluation & Tuning**: Use metrics like accuracy and F1-score; optimize hyperparameters for best performance.
3. **Cross-Validation**: Ensure model generalization to avoid overfitting.

**Outcome**: A highly accurate deep learning model for CWR identification.

**Phase 4: Real-time Identification System Development**

**Objective**: Create deployable systems for field use.

1. **Model Optimization**: Use quantization and pruning for mobile and cloud deployment.
2. **Mobile & Cloud Integration**: Develop a mobile app and cloud service for real-time identification.
3. **Field Testing**: Validate the system in diverse conditions and refine based on user feedback.

**Outcome**: A real-time, field-ready system supporting conservation and research.

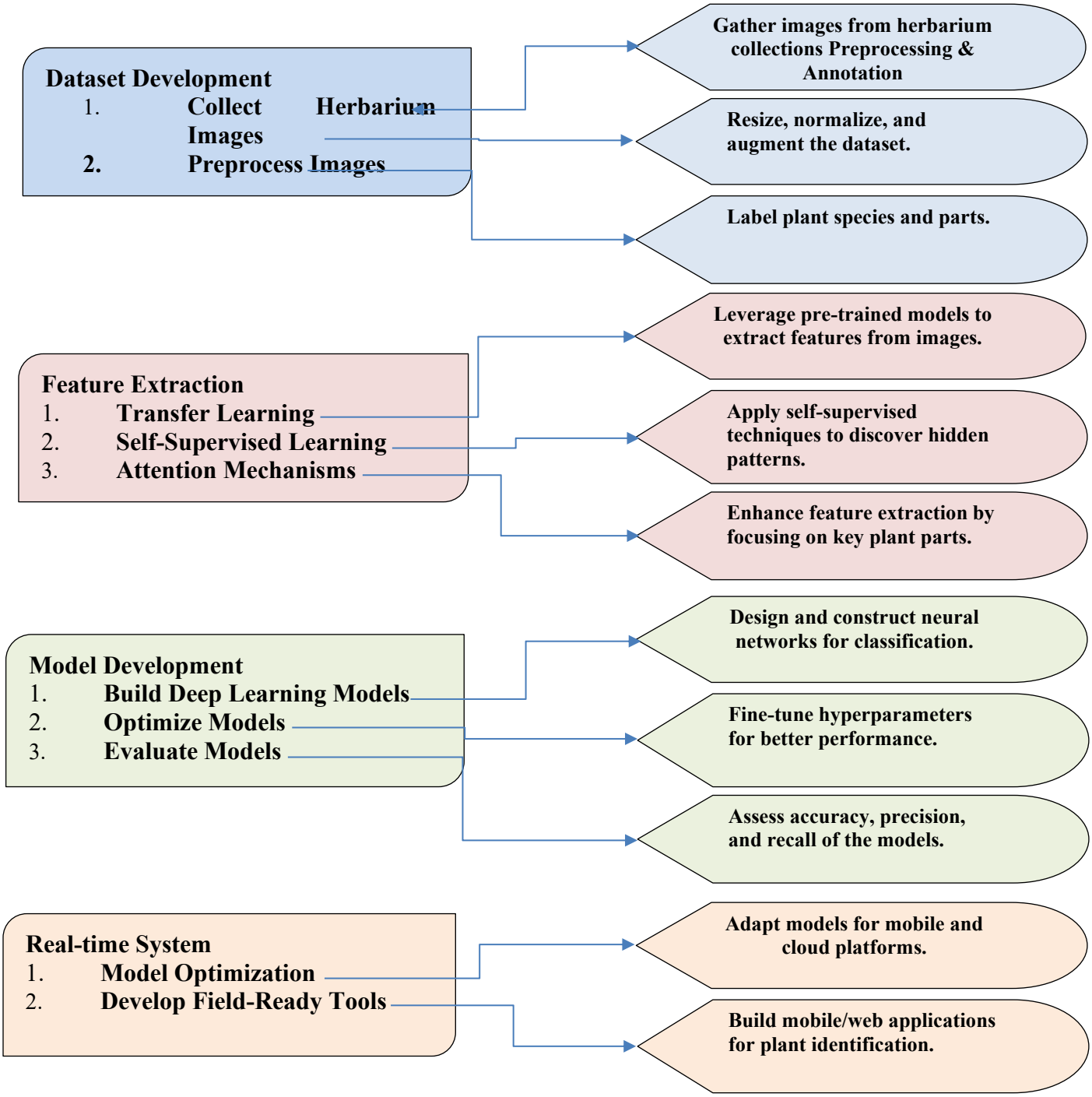
Dataset Development
1. Collect Herbarium Images
2. Preprocess Images
Gather images from herbarium collections Preprocessing & Annotation
Resize, normalize, and augment the dataset.
Label plant species and parts.
Feature Extraction
1. Transfer Learning
2. Self-Supervised Learning
3. Attention Mechanisms
Leverage pre-trained models to extract features from images.
Apply self-supervised techniques to discover hidden patterns.
Enhance feature extraction by focusing on key plant parts.
Model Development
1. Build Deep Learning Models
2. Optimize Models
3. Evaluate Models
Design and construct neural networks for classification.
Fine-tune hyperparameters for better performance.
Assess accuracy, precision, and recall of the models.
Real-time System
1. Model Optimization
2. Develop Field-Ready Tools
Adapt models for mobile and cloud platforms.
Build mobile/web applications for plant identification.

Figure 3: Proposed Methodology

## 7. Expected Outcome of the Research

The main finding of this synthesis is that advanced deep learning models are nearing the level of expertise demonstrated by humans. The proposed work aims to identify both whole plants from the CWR Herbarium and live plants, benefiting future researchers, taxonomists, botanists, and students. This study will also contribute novel CWR species to the National Herbarium of Cultivated Plants, ICAR-NBPGR, New Delhi, which will help train the model using datasets from various periods. Once the model is successfully developed, it will be implemented in a user-friendly application for easier access and utilization.

## References


[1] "Role of Agriculture in Economic Development in India." Accessed: Oct. 04, 2024. [Online]. Available: https://www.economicsdiscussion.net/agricultural-economics/role-of-agriculture-in-economic-development-india-economics/26748

[2] "Role of Agriculture in Indian Economy." Accessed: Oct. 04, 2024. [Online]. Available: https://www.agrifarming.in/role-of-agriculture-in-indian-economy

[3] V. Geethalakshmi *et al.*, "Climate Change and Its Impact on Agriculture in India," *Agrometeorological Applications for Climate Resilient Agriculture*, pp. 261–282, 2024, doi: 10.1007/978-3-031-51083-0_13.

[4] A. Kapazoglou *et al.*, "Crop Wild Relatives: A Valuable Source of Tolerance to Various Abiotic Stresses," *Plants 2023, Vol. 12, Page 328*, vol. 12, no. 2, p. 328, Jan. 2023, doi: 10.3390/PLANTS12020328.

[5] H. Vincent *et al.*, "Modeling of crop wild relative species identifies areas globally for in situ conservation," *Communications Biology 2019 2:1*, vol. 2, no. 1, pp. 1–8, Apr. 2019, doi: 10.1038/s42003-019-0372-z.

[6] A. Bohra *et al.*, "Reap the crop wild relatives for breeding future crops," *Trends Biotechnol*, vol. 40, no. 4, pp. 412–431, Apr. 2022, doi: 10.1016/J.TIBTECH.2021.08.009.

[7] "ICRISAT." Accessed: Oct. 04, 2024. [Online]. Available: https://www.icrisat.org/

[8] "ICARDA | Science for resilient livelihoods in dry areas." Accessed: Oct. 04, 2024. [Online]. Available: https://www.icarda.org/

[9] "CIMMYT." Accessed: Oct. 04, 2024. [Online]. Available: https://www.cimmyt.org/

[10] "CGIAR: Science for humanity's greatest challenges." Accessed: Oct. 04, 2024. [Online]. Available: https://www.cgiar.org/

[11] A. Pandey, K. Pradheep, R. Gupta, and S. P. Ahlawat, "National Herbarium of Cultivated Plants: A Resource for Study of Crop Genepools," *Indian Journal of Plant Genetic Resources*, vol. 30, no. 03, pp. 246–252, Nov. 2017, doi: 10.5958/0976-1926.2017.00031.6.

[12] "(PDF) Crop Wild Relatives in India: Prioritisation, Collection and Conservation." Accessed: Oct. 04, 2024. [Online]. Available: https://www.researchgate.net/publication/362545581_Crop_Wild_Relatives_in_India_Prioritisation_Collection_and_Conservation

[13] "National Herbarium Cultivated Plant." Accessed: Oct. 04, 2024. [Online]. Available: http://www.nbpgr.ernet.in:8080/nhcp/

[14] "Crop Wild Relatives: The importance of Crop Wild Relatives." Accessed: Oct. 04, 2024. [Online]. Available: http://www.cropwildrelatives.org/resources/in-situ-conservation-manual/elearning-modules/elearning/the-importance-of-crop-wild-relatives

[15] H. Yalcin and S. Razavi, "Plant classification using convolutional neural networks," *2016 5th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2016*, Sep. 2016, doi: 10.1109/AGRO-GEOINFORMATICS.2016.7577698.

[16] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, "Deep-plant: Plant identification with convolutional neural networks," *Proceedings - International Conference on Image Processing, ICIP*, vol. 2015-December, pp. 452–456, Dec. 2015, doi: 10.1109/ICIP.2015.7350839.

[17] M. Shirai *et al.*, "Development of a system for the automated identification of herbarium specimens with high accuracy," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-11450-y.

[18] A. Triki, B. Bouaziz, J. Gaikwad, and W. Mahdi, "Deep eaf: Mask R-CNN based leaf detection and segmentation from digitized herbarium specimen images," *Pattern Recognit Lett*, vol. 150, pp. 76–83, Oct. 2021, doi: 10.1016/j.patrec.2021.07.003.

[19] W. N. Weaver, J. Ng, and R. G. Laport, "LeafMachine: Using machine learning to automate leaf trait extraction from digitized herbarium specimens," *Appl Plant Sci*, vol. 8, no. 6, Jun. 2020, doi: 10.1002/aps3.11367.

[20] S. Younis, M. Schmidt, C. Weiland, S. Dressler, B. Seeger, and T. Hickler, "Detection and annotation of plant organs from digitised herbarium scans using deep learning," *Biodivers Data J*, vol. 8, pp. 1–18, 2020, doi: 10.3897/BDJ.8.E57090.

[21] S. Chulif, S. H. Lee, Y. L. Chang, and K. C. Chai, "A machine learning approach for cross-domain plant identification using herbarium specimens," *Neural Comput Appl*, vol. 35, no. 8, pp. 5963–5985, Mar. 2023, doi: 10.1007/s00521-022-07951-6.

[22] N. Iwanycki Ahlstrand, H. M. Elvery, and R. B. Primack, "Grass flowering times determined using herbarium specimens for modeling grass pollen under a warming climate," *Science of the Total Environment*, vol. 885, Aug. 2023, doi: 10.1016/j.scitotenv.2023.163824.
[23] B. R. Hussein, O. A. Malik, W. H. Ong, and J. W. F. Slik, "Reconstruction of damaged herbarium leaves using deep learning techniques for improving classification accuracy," *Ecol Inform*, vol. 61, Mar. 2021, doi: 10.1016/j.ecoinf.2021.101243.
[24] D. P. Little *et al.*, "An algorithm competition for automatic species identification from herbarium specimens," *Appl Plant Sci*, vol. 8, no. 6, Jun. 2020, doi: 10.1002/aps3.11365.
[25] T. Ott, C. Palm, R. Vogt, and C. Oberprieler, "GinJinn: An object-detection pipeline for automated feature extraction from herbarium specimens," *Appl Plant Sci*, vol. 8, no. 6, Jun. 2020, doi: 10.1002/aps3.11351.
[26] K. S. Jye, S. Manickam, S. Malek, M. Mosleh, and S. K. Dhillon, "Automated plant identification using artificial neural network and support vector machine," *Front Life Sci*, vol. 10, no. 1, pp. 98–107, 2017, doi: 10.1080/21553769.2017.1412361.
[27] J. Wäldchen and P. Mäder, "Machine learning for image based species identification," Nov. 01, 2018, *British Ecological Society*. doi: 10.1111/2041-210X.13075.
[28] A. B. Ballo, D. Mamadou, K. J. Ayikpa, K. Yao, E. A. A. Ablan, and K. F. Kouame, "Automatic Identification of Ivorian Plants from Herbarium Specimens using Deep Learning," *International Journal of Emerging Technology and Advanced Engineering*, vol. 12, no. 5, pp. 56–66, May 2022, doi: 10.46338/ijetae0522_07.
[29] M. Opała-Owczarek and P. Owczarek, "Herbarium records in Arctic dwarf shrub dendrochronology: Methodological approach and perspectives," *Dendrochronologia (Verona)*, vol. 80, Aug. 2023, doi: 10.1016/j.dendro.2023.126102.
[30] I. Latin and A. Transactions, "Senescence Reversion in Plant Images using Perception and Unpaired Data," 2022.
[31] B. R. Hussein, O. A. Malik, W. H. Ong, and J. W. F. Slik, "Applications of computer vision and machine learning techniques for digitized herbarium specimens: A systematic literature review," Jul. 01, 2022, *Elsevier B.V.* doi: 10.1016/j.ecoinf.2022.101641.
[32] J. H. Tibbetts, "The Frontiers of Artificial Intelligence," *Bioscience*, vol. 68, no. 1, pp. 5–10, Jan. 2018, doi: 10.1093/biosci/bix136.
[33] S. Chulif and Y. L. Chang, "Improved Herbarium-Field Triplet Network for Cross-Domain Plant Identification: NEUON Submission to LifeCLEF 2021 Plant," 2021. [Online]. Available: http://ceur-ws.org
[34] A. Y. Kajihara, D. Bertolini, and A. L. Schwerz, "Identification of herbarium specimens: a case study with Piperaceae Giseke family," *International Conference on Systems, Signals, and Image Processing*, vol. 2022-June, 2022, doi: 10.1109/IWSSIP55020.2022.9854444.
[35] A. Triki, B. Bouaziz, W. Mahdi, and J. Gaikwad, "Objects Detection from Digitized Herbarium Specimen based on Improved YOLO V3," *VISIGRAPP 2020 - Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 4, pp. 523–529, 2020, doi: 10.5220/0009170005230529.
[36] Available: http://www.nbpgr.ernet.in:8080/nhcp/zoo.aspx?imageIdUrl=1770
[37] Available: http://www.nbpgr.ernet.in:8080/nhcp/zoo.aspx?imageIdUrl=1350

MANAV RACHNA
Vidyanatariksha

**Dr. O.P. Bhalla Central Library**
**Manav Rachna International Institute of Research and Studies**
*(A deemed-to-be-university under section 3 of UGC Act 1956)*

## SIMILARITY VERIFICATION CERTIFICATE

Title: DESIGNING DEEP LEARNING MODEL FOR AUTOMATIC IDENTIFICATION OF CROP WILD RELATIVES

Researcher: ANKUR TOMAR
Registration Number: 22/PHd/S/043

Supervisor: DR. SURESH KUMAR
Department: SET, CSE

This is to certify that the above mentioned Short Synopsis/Thesis was scanned for similarity index through Turnitin Software on 25/03/2026 with submission ID 135166528 and the similarity index was found to be 09 %. The complete report running into 23 Pages, 6288 Word count with File Size 3.2 KB/MB is being forwarded to the Research Supervisor/DRC/DRB for review.

Date: 25/03/2026

Dr. Rajesh Kumar
University Librarian
Dr. O.P. Bhalla Central Library
Manav Rachna International
Institute of Research and Studies
Faridabad

The complete report of the above Short Synopsis/Thesis has been reviewed and the similarity index is below the accepted norms as per the University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations 2018. The Short Synopsis/Thesis may be considered for further processing

Research Scholar

Research Supervisor

Excluding matches
less than 14 words

Dr. Rajesh Kumar
University Librarian
Dr. O.P. Sinha Central Library
Manav Rachna International

# Ankur Tomar

## DESIGNING DEEP LEARNING MODEL FOR AUTOMATIC IDENTIFICATION OF CROP WILD RELATIVES

Quick Submit

Quick Submit

Manav Rachna International University

### Document Details

Submission ID
trn:oid:::1:3516659528

Submission Date
Mar 25, 2026, 9:17 AM GMT+5:30

Download Date
Mar 25, 2026, 9:21 AM GMT+5:30

File Name
G_MODEL_FOR_AUTOMATIC_IDENTIFICATION_OF_CROP_WILD_RELATIVES.docx

File Size
3.2 MB

23 Pages

6,288 Words

39,354 Characters

# 9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

## Filtered from the Report

- Bibliography
- Small Matches (less than 12 words)

Excluding matches less than 14 words

Dr. Rajesh Kumar
University Librarian
Dr. O.P. Bhatia Central Library
Manav Rachna International
Institute of Research and

## Match Groups

- **21** Not Cited or Quoted **8%**
  Matches with neither in-text citation nor quotation marks
- **5** Missing Quotations **2%**
  Matches that are still very similar to source material
- **0** Missing Citation **0%**
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted **0%**
  Matches with in-text citation present, but no quotation marks

## Top Sources

4% Internet sources
8% Publications
2% Submitted works (Student Papers)

## Integrity Flags

**1 Integrity Flag for Review**

**Hidden Text**
37 suspect characters on 1 page
Text is altered to blend into the white background of the document.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

## Match Groups

- **21** Not Cited or Quoted 8%
  Matches with neither in-text citation nor quotation marks
- **5** Missing Quotations 2%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

## Top Sources

- 4% Internet sources
- 8% Publications
- 2% Submitted works (Student Papers)

Excluding matches less than 14 words

Dr. Rajesh Kumar
University Librarian
Dr. O.P. Bhalla Central Library
Manav Rachna International
Institute of Research & Studies
Faridabad

## Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

| | | |
|---|---|---|
| 1 | Publication | |
| | Ankur Tomar, Suresh Kumar, Kuldeep Tripathi. "Review of herbarium plant identi... | 4% |
| 2 | Publication | |
| | Ankur Tomar, Suresh Kumar, Kuldeep Tripathi. "Comparative analysis of AlexNet ... | 2% |
| 3 | Internet | |
| | www.researchgate.net | <1% |
| 4 | Internet | |
| | ijece.iaescore.com | <1% |
| 5 | Internet | |
| | www.nbpgr.ernet.in:8080 | <1% |
| 6 | Internet | |
| | escholarship.org | <1% |
| 7 | Internet | |
| | www.mdpi.com | <1% |
| 8 | Student papers | |
| | Liverpool John Moores University | <1% |
| 9 | Internet | |
| | phytopatholres.biomedcentral.com | <1% |
| 10 | Publication | |
| | Faisal Firdous, Deepak Gupta, Hemant Sood. "AI-Based Smart Identification of Me... | <1% |

11 Internet

api.deepai.org <1%

12 Student papers

University of the Philippines - Manila <1%

13 Publication

William N. Weaver, Julienne Ng, Robert G. Laport. "LeafMachine: Using machine l... <1%

# DESIGNING DEEP LEARNING MODEL FOR AUTOMATIC IDENTIFICATION OF CROP WILD RELATIVES

## SHORT SYNOPSIS

by

**Ankur Tomar**
Registration Number: 22/PHD/S/043

*Under the Supervision of*

Supervisor
Dr. Suresh Kumar
Professor,
CSE Department, SET

Co-Supervisor
Dr. Kuldeep Tripathi
Senior Scientist,
ICAR-NBPGR, New Delhi

MANAV RACHNA
||vidyanatariksha||

**Department of Computer Science & Engineering**
**School of Engineering & Technology**
**Manav Rachna International Institute of Research & Studies**
**Sector 43, Faridabad, India.**

## Abstract

Herbarium specimens contain a vast wealth of genetic resources. Among these resources are the Crop Wild Relatives (CWRs) which play a crucial role in improving crops and preserving genetic diversity. Identifying CWRs automatically from live plant and herbarium images is a complex challenge due to the variations in plant shapes and the need for accurate species recognition. Convolutional Neural Networks (CNNs) have proven to be highly effective tools for classifying images, offering promising possibilities for automating CWR identification. This concise summary delves into CNN models specifically designed for plant identification across various formats, including herbarium specimens. The underlying design principles are explored, performance evaluated, and practical applications highlighted in safeguarding biodiversity and advancing agricultural research.

In recent years, researchers have made significant advancements in the development of Convolutional Neural Networks (CNNs) tailored for CWR identification in herbarium